The visual representation of perception of American English vowels by university-level Korean listeners: a multidimensional scaling-based model[*] [**]

Soonhyun Hong
(Inha University)

Hong, Soonhyun. 2011. The visual representation of perception of American English vowels by university-level Korean listeners: a multidimensional scaling-based model. *Studies in Phonetics, Phonology and Morphology* 17.1, 121-144. This paper tries to represent Korean listeners' relative perceptual similarities between nine American English vowels on a two-dimensional perceptual space through a multidimensional scaling analysis of the confusion matrices from identification tests. It is demonstrated that the derived perceptual vowel space through a multidimensional scaling analysis is visually quite similar to the acoustic vowel map with F1 and F2 values in the literature. In the case of Korean listeners' American English vowel perception, as the perceptual vowel map represents perceptually similar vowels close to each other on a two-dimensional map, it will become easier to visually interpret which vowel pair a Korean listener has perceptual difficulties with. An optimal model with a multidimensional scaling analysis is also proposed to drastically reduce stress by splitting the confusion matrices into front and back vowel groups and conducting multidimensional scaling analyses separately. The split matrix model can explain the data better than the single matrix model. **(Inha University)**

Keywords: English vowel perception by Korean listeners, Multidimensional Scaling, perceptual vowel space, MDS, vowel features

1. Introduction

In the literature on L2 learning, native listeners of a language tend to perceive non-native sounds according to the phonetic and phonological patterns of L1 categories (Best 1995, Flege 1988, 1995). Some cross-language studies adopted identification or discrimination tasks to evaluate listeners' perception on L2 vowels. For example, Japanese learners of American English (AE) had difficulty distinguishing between mid vowels and between low vowels (Best 1995, Lambacher et al. 2000, Yamada et al. 1995). The reason for this is that Japanese does not have two distinctive mid and low vowels and Japanese learners have more difficulty perceiving AE mid and low vowels (Best 1995).

A lot of different studies have reported L2 vowel perception by speakers of different languages (Stevens et al. 1969, Terbeek 1977, Flege et al.

[*] This work was supported by the National Research Foundation of Korea Grant funded by the Korean government (NRF-327-2009-1-A00288).

[**] I would like to thank three anonymous reviewers for their insightful comments and suggestions on this paper.

1994). These studies suggest that Korean listeners' perception of AE vowels may not pattern together with that of AE listeners'. However, it is not easy to characterize or explain the perceptual inter-relationship among AE vowels which is perceived by Korean listeners of AE vowels. For example, one AE vowel may be perceived more similar to another vowel by Korean listeners than to the other vowels. Another vowel may be perceptually similar to yet another vowel to Korean listeners. These more confusable similar pairs of AE vowels by Korean listeners are the pairs of AE vowels which are not phonemic in Korean. For example, AE vowel pairs, [iy] vs. [i], and [e] vs. [æ], and [uw] vs. [u], are not distinctive in Korean and the two members of each pair are very difficult for Korean listeners to distinguish perceptually.

The problem is that, as is to be shown in the AE vowel identification experiment in this paper, the perceptual confusability varies between Korean listeners as well as among AE vowel pairs within Korean listeners. Namely, different Korean listeners show quite different perceptual abilities between perceptually similar AE vowels in those AE vowel pairs. This perceptual variability among Korean listeners is actually predicted considering the fact that they have different backgrounds of English study and different English acquisition capabilities. However, little study has yet focused on the evaluation of relative perceptual confusability among AE vowels within Korean listeners. In fact, the perceptual confusability among the three AE vowel pairs varies greatly for different Korean listeners. Namely, some Korean listeners felt perceptually more difficult in the [iy] and [i] pair than in the [e] and [æ] and the [uw] and [u] pairs while some others suffer more perceptual difficulties with the [uw] and [u] pair than with the other pairs. If we would like to help Korean listeners enhance their own AE vowel perceptional abilities, we need to diagnose their perceptual abilities as to which AE vowel pairs are perceptually problematic or not. This way, we can let them know their own perceptual problems and might devise a perceptual training program on AE vowels (Nishi and Kewley-Port 2007). As for a training program of AE vowels for Korean listeners, the training procedures are usually time-consuming and tedious, thus the number of training vowel stimuli should be cut down as many as possible. An adaptive AE vowel minimal pair training program (Hong 2009 for AE fricatives), for example, could be devised, which provides Korean listener with more vowel stimuli of perceptually difficult vowel pairs during the training, but with less number of vowel stimuli of perceptually easy vowel pairs.

Hence, the within-subject AE vowel perceptual abilities should be considered seriously: Which vowel pairs are perceptually more difficult for a given Korean listener? For this purpose, a multidimensional scaling (MDS) analysis is to be conducted, which graphically shows which two vowels are perceptually more similar to each other than the other vowels. The similarity or dissimilarity between vowels is represented on a 2- or

3-dimensional perceptual space through an MDS analysis of the confusion matrices from an AE vowel identification test. Namely, two perceptually similar vowels will be represented close to each other on the perceptual map whereas perceptually different vowels will be represented as points distant from each other. Such information will help Korean listeners see their own perceptual vowel map, which turns out to be very similar to their acoustic vowel map with F1 and F2 values. It will be suggested that the MDS-based perceptual vowel map might be one potential clue to solve the long-lasting puzzle to define the relationship between perception and production.

On the production side, how Korean speaker produce AE vowels can be visually represented in an acoustic vowel space through coordinates of F1 and F2, as in Peterson and Barney (hereafter, P&B) (1952:182). On the perception side, however, less is known as to how AE vowels are represented visually in a perceptual vowel space.

A move to visually represent Korean listeners' perception of AE vowels is found in Hong (2007a). In the experiment, naturally spoken CV stimuli were prepared with the last 15 percent of the vowel removed to avoid a lexical effect. Then 20 Korean listeners were forced to identify AE vowels after they heard the stimuli. The results were fed onto a Hierarchical Cluster Analysis. The resulting dendrogram visually showed that AE pairs /uw/ vs. /u/, /e/ vs. /æ/, and /iy/ vs. /i/ were most frequently confused from each other. In the dendrogram in figure 1, the more confusable the pair is, the more leftward the pair node is formed. Note that the accuracy rates are shown in the parenthesis. Note that [V] refers to [ʌ].

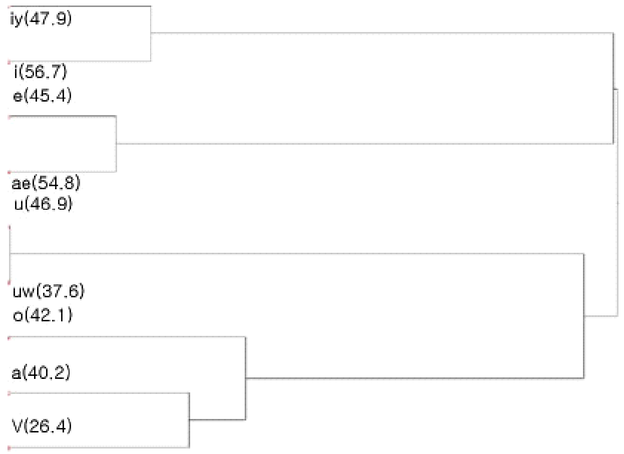


Figure 1. Dendrogram for 20 Korean learners of English with accuracies in parenthesis

It was found that university-level Korean listeners had the most identification difficulty between AE /u/ and /uw/. They also had identification difficulty in pairs /e/ vs. /æ/ and /iy/ vs. /i/.

On the other hand, the following dendrogram is for a native AE listener:

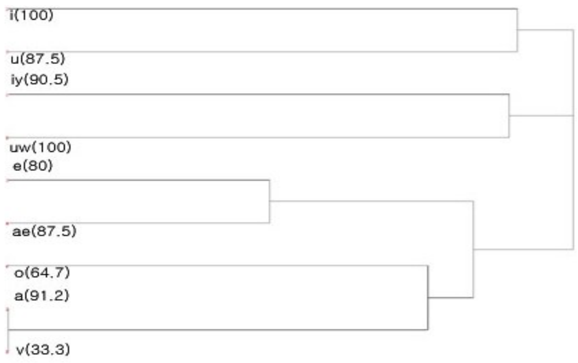


Figure 2. Dendrogram for a native monolingual American English male from Pennsylvania

This native male AE listener had no problems in perception of AE vowels except that he had difficulties differentiating between [a] and [ʌ].

The dendrograms for AE vowel confusion by Korean listeners may be a potentially good tool to show what kind of confusion problems Korean listeners have in AE vowel perception. However, dendrograms are very difficult to interpret. The horizontal scale is formed relative to reciprocal perception similarities between vowels and hence becomes another variable changing from a dendrogram to another.

The multivariate statistical MDS can be a better analysis for this case than the cluster analysis. It can visually model the perceptual vowel space with a structure of relatively few dimensions, based on the perceptual judgments by the listeners (Singh and Woods 1970, Carroll and Jang 1970, Kruskal and Wish 1978, Terbeek 1977).

In an effort to represent vowel perception on a perceptual space, it was found in the literature that the dimensions resulting from an MDS analysis can be correlated to some phonological features, suggesting that phonological features play significant roles in vowel perception (Singh and Woods 1970, Shepard 1972, Terbeek 1977, Fox 1983, Fox et al. 1995).

The purpose of the present work is to visually represent Korean listeners' perception of AE vowels on a perceptual space. For this purpose, an identification test was waged to evaluate 34 Korean listener's perception of AE vowels and the resulting confusion matrices were fed onto an MDS analysis for visualization and interpretation.

Two different identification experiments were waged. The purpose of the first experiment was to demonstrate how to represent Korean listeners' and native listeners' perception of AE vowels on a 2-dimensional perceptual space. The CVC stimuli with all possible consonantal environments were selected and the target vowels were manipulated to trigger more errors in the perception of AE vowels by Korean and native listeners. Since the MDS analysis in this paper crucially uses relative perceptual errors between vowels by listeners, perceptual vowel space would not be derived for the native listeners without their perceptual errors. Thus the purpose of the first experiment was to show that MDS is an appropriate analysis to represent perceptual vowel spaces from Korean listeners and native listeners. On the other hand, the purpose of the second experiment was to build the most optimal model to represent perceptual vowel space for only Korean listeners both on the whole and on the individual basis. Thus only hVd stimuli were used in the second experiment[1].

2. Experiment I: Identification test with manipulation

2.1 Subjects

The subjects were 34 university-level Korean students (7 males and 27 females) ranging in age from 20 to 25 years old, who were taking English phonetics in a Korean university. All of them had at least 6 years of prior English instruction at the middle and high school levels. As this experiment was waged at the beginning of the semester, they had no phonetic knowledge of AE vowels. Two native AE listeners (2 males) of 60 and 24 years old from eastern part of the US also participated in the experiment for reference. None of the subjects had any reported history of speech or hearing problems.

2.2 Stimuli

The objective of the experiment was to force the Korean listeners to identify AE vowels in response to the presentation of AE CV speech stimuli. For this purpose, consonant-vowel matrix was formed with the rows of 18 English consonants /p, b, t, d, k, g, f, v, th(ɵ), dh(ð), s, z, sh(ʃ), ch(tʃ), dz(dʒ), h, l, r/[2] for AE consonants and with the columns of 9 primary-stressed target AE vowels /iy(i), i(ɪ), e(ɛ), æ, a, o(ɔ), u(ʊ), uw(u), X(ʌ)/[3]. As for the consonants, /ʒ/ and /ŋ/ were not considered since they do not appear in word-initial position in English. The other nasals were also excluded from the matrix to cut down the number of the stimuli. In the

[1] The second experiment is not appropriate for analyzing native listeners' perception of AE vowels through MDS, since they rarely made perceptual errors in the perception of hVd stimuli.

[2] Due to incompatibility with the statistical package, /p, b, t, d, k, g, f, v, th, dh, s, z, sh, ch, dz, h, l, r/ will be used in this paper.

[3] Due to incompatibility with the statistical package, /iy, i, e, æ, a, o, u, uw, X/ will be used in this paper.

matrix, each cell was occupied by a one- or two-syllable English word beginning with the target CV stimulus. 16 empty cells arose due to English phonotactics or unavailability of spoken word samples. The spoken samples for the words in the matrix were extracted from Yahoo Dictionary and were resampled at a 48 kHz sampling rate. Each spoken sample was verified through careful listening by the author and about 15% of the last portion of the first vowel in the #CV... sample was removed along with following segments[4] through the author's careful listening and through referring to its corresponding waveform and spectrogram with Wavesurfer 1.8.1. It was assumed that this process would make sure that the quality of the target vowel for identification be as invariant as possible by the author's judgment. Furthermore, the identification of the target vowel in the stimulus CV speech sample, would be less affected by potential lexical effects since it would become more difficult for the subject to guess which word the target CV stimulus was extracted from. Such a trimming process might not affect the vowel perception significantly under the assumption that the surface forms of auditory stimuli are retained in memory in the form of exemplars and can be retrieved upon request. Strange, Jenkins and Johnson (1983) demonstrated that the vowels in spoken bVb syllables were modified to generate 7 modified syllable conditions in which different parts of the digitized waveform of the syllables in question were deleted and the temporal relationships of the remaining parts were manipulated. The identification results of vowels by untrained listeners showed that dynamic spectral information, contained in initial and final transitions taken together, was sufficient for accurate identification of vowels even when vowel nuclei were attenuated to silence. As for the vowel stimuli in the present experiment, the initial and the proportionally variable center were retained for vowel identification.

The spoken samples of those words were spoken by an unidentifiable number of male and female speakers with their voices interspersed in 146 speech tokens.

Table 1. CV matrix of stimuli

C V	iy	i	e	æ	a	o	u	uw	X	total
p	peace	pin	pen	past	pot	pour	push	pool	pup	9
b	beach	bin	bench	bath	bobby	ball	book	boom	bub	9
t	team	tin	tent	taxi	top	talk	took	tool	touch	9
d	deep	dish	desk	dance	dot	dog		deuce	dust	8
k	keep	kick	keg	cap	common	corer	cook		cub	8

[4] As a reviewer pointed out, the manipulated stimuli in this experiment could be replaced by synthesized stimuli. There might arise some difference in the results depending on the characteristics of the types of the stimuli, which is a pending question for a further research.

g	geese	gift	guess	gather	goddess	gawk	good	goose	gun	9
f	feel	fig	fen	fan	fox	four	full	fool	fun	9
v	veal	viz	vet	valley	volume	vault		voodoo	vulgar	8
th	thief	think	theft	thank		thorn			thumb	6
dh	these	this	them	that					thus	5
s	seed	sit	send	sad	sop	salt	soot	soup	sub	9
z	zeal	zip	zest	zapper				zoom		5
sh	sheep	shift	shelf	shadow	shop	shawl	should	shoe	shovel	9
ch	cheese	chick	check	chance	chop	chalk		choose	chum	8
dz	jeep	gin	jet	jacket	jar	jaw		juice	jug	8
h	heat	hit	heck	habit	hot	horse	hook	hoodoo	hut	9
l	leak	lick	let	lack	lock	law	look	loose	lush	9
r	reap	rick	rest	rack	rock	wrong	rook	rule	rush	9
									grand total	146

2.3 Procedure

The AE consonant and vowel perception experiment was done through the presentation of the randomized 146 CV token which were repeated two times, totaling 292 presentations (=146*2) for each participant. The number of total stimulus presentations to all 34 participants was 9928 (=292 presentations*34 listeners).

A brief introduction to 18 AE consonants and 9 AE vowels and their respective transcriptions was given to the listeners beforehand. Then a testing tool based on Alvin (Hillenbrand et al. 2005), provided a computer screen showing the 18 consonant icons with example words on the left half of the screen and 9 vowel icons with example words on the right half of the screen, at the same time. As listeners were forced to identify both the consonant and the vowel, after each presentation of the stimuli, listeners did not know that the target segments are vowels. Each of the consonant icons was filled with an IPA symbol and an example word containing the consonant in word-initial position. The example words are *boy, desk, girl, pot, ten, kid, right, lip, video, this, zero, fan, think sun, ship, home, jeep,* and *chair*. Each of the vowel icons was filled with an IPA symbol and an example word containing the vowel: *heed*, *hid*, *head*, *had*, *hod*, *hawed*, *hood*, *who'd*, and *hud*. No problem was reported during the experiment since subjects were already familiar with the example words and IPA symbols.

Each listener heard randomized CV stimuli presented via a PC over a headphone, and was forced to click on one of the 18 consonant icons and then one of the 9 vowel icons after each CV presentation. The sound volume could be adjustable by listeners. After the two clicks, one for the consonant and the other for vowel identification, there was a pause of 400ms before the next stimulus was presented. Without clicking, the next stimulus was not presented. When a listener made an error click(s), s/he could go back and make readjustment clicks after listening to the previous presentation again. This experiment setup allowed participants to proceed at their own comfortable pace. Furthermore, s/he could listen to each presentation repeated up to 3 times. The identification experiment for each listener took about less than 25 minutes on average.

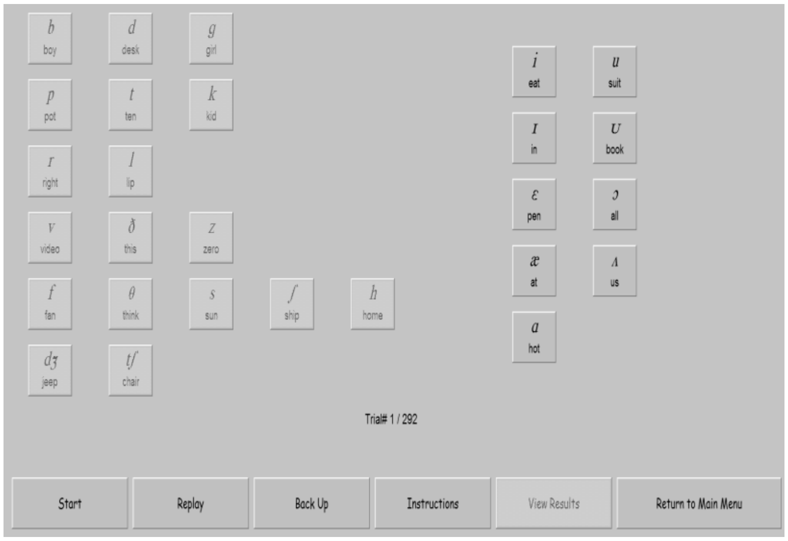


Figure 3. Computer screen for CV identification test

2.4 Results

The mean accuracy rates for AE consonants and vowels for all listeners are 73.88% and 50.76%, respectively.

Table 2. Consonant and vowel accuracy rates

	Mean	s.d.	N
C accuracy	73.88	7.307	34
V accuracy	50.76	6.068	34

The mean accuracy rate for the perception of AE vowels by all the Korean listeners was 50.52%. The following 95% CI errorbars show that /a/ is the easiest for Korean listeners. Front /e/ and back /u, uw, X/ were more difficult than the other vowels. /a/ was the easiest for Korean listeners.

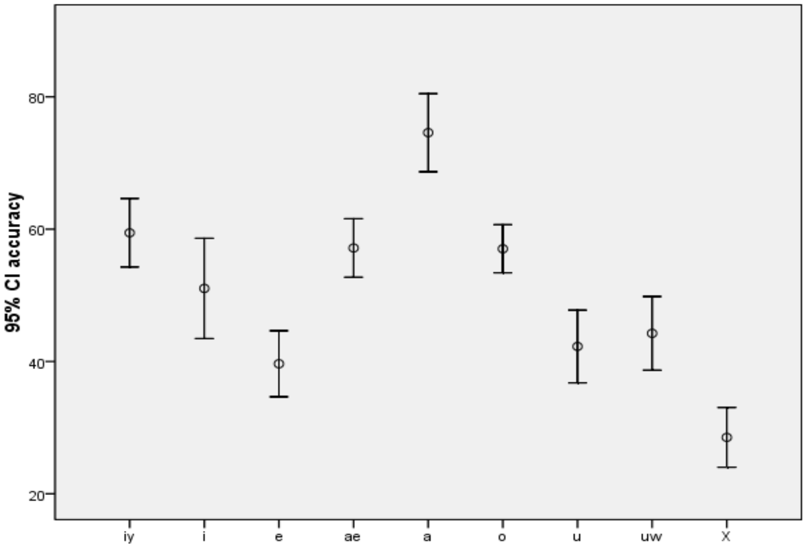


Figure 4. 95% CI errorbars for mean accuracy rates of perception of AE vowels by Korean listeners

Unfortunately, this graph above provided no information about Korean listeners' relative perceptual difficulties between individual vowels. In what follows, we will mainly focus on the analysis and discussions of AE vowel perception and perceptual interference between vowels by Korean listeners through an MDS analysis.

3. Discussions: An MDS analysis of relative AE vowel perception

3.1 Introduction to MDS

Before we conduct an MDS analysis, we will provide a brief introduction to how an MDS analysis works. The analysis can provide a visual representation of the pattern of proximities (i.e., similarities or distances) among a set of objects. A typical example is found in the matrix of distances among American cities. Given only the local distances between cities, an MDS analysis can derive a 2-dimensional map which visually represents all the points of cities. The relationships between input distances among city points are calculated, using the Euclidean distance formula below:

$$d_{ij} = \sqrt{\sum_{k=1}^{n} (X_{ik} - X_{jk})^2}$$

Table 3. The matrix of distances among US cities

	Bost	NYC	DC	Miam	Chic	Seat	SF	LA	Denv
Bost	0	206	429	1504	963	2976	3095	2979	1949
NYC	206	0	233	1308	802	2815	2934	2786	1771
DC	429	233	0	1075	671	2684	2799	2631	1616
Miam	1504	1308	1075	0	1329	3273	3053	2687	2037
Chic	963	802	671	1329	0	2013	2142	2054	996
Seat	2976	2815	2684	3273	2013	0	808	1131	1307
SF	3095	2934	2799	3053	2142	808	0	379	1235
LA	2979	2786	2631	2687	2054	1131	379	0	1059
Denv	1949	1771	1616	2037	996	1307	1235	1059	0

The symmetric matrix above shows the distance in miles between US cities in each cell. The resulting map from an MDS analysis represents the city points on a 2-dimensional perceptual space:

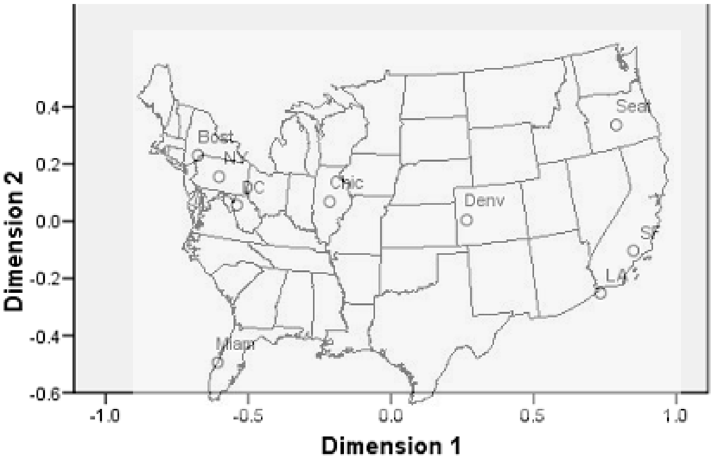


Figure 5. The perceptual space of US cities is overlapped on with a mirror image of a US map

The visual proximities representation above is the exact mirror image of a US map representing all the major city points, as shown with a blank US map overlapped on. The resulting map shows that the derived US city points are well fitted onto the blank mirror image US map. The smaller the inter-city proximity, the closer the two city points on the perceptual space. If the input data were similarities, the smaller the input similarity between objects, the farther apart two object points would be on the perceptual space.

Since the orientation of the perceptual space is arbitrarily represented by an MDS analysis, the resulting representation of distances between US cities should be interpreted such that east is left and west is right, as shown the derived map for US cities. All that matters is which point is close to which others.

3.2 Comparison between perceptual and acoustic vowel spaces for AE talkers vs. listeners

PROXSCAL MDS seeks a perceptual-spatial representation for listeners' AE vowel identification errors by converting the confusions (incorrect vowel identification cases) into identification similarities. It can provide a representation of the inter-vowel relationships in a low-dimensional perceptual vowel space through the use of the proximities between the vowels. It minimizes the squared deviations between the possibly transformed inter-vowel proximities and their Euclidean distances, representing the relative distance between the vowels in the low-dimensional perceptual space. The smaller the distances are between vowel stimulus, the more confusable the vowels are; the larger the distances, the more readily distinguished the vowels are from one another.

Before an MDS analysis of the resulting data from AE listeners was conducted, two tasks were done: building an MDS-derived perceptual vowel map for the two AE listeners' data and plotting F1 and F2 values of vowels on an acoustic map. Then the two maps were compared side by side.

First, confusion matrices were built from the results of the identification test for AE listeners. The following confusion matrix is one sample for one AE listener:

Table 4. AE vowel confusion matrix with accuracy rates for one AE listener

listener1	iy	i	e	æ	a	o	u	uw	X
iy	80.56	13.89	0	0	0	0	2.78	2.78	0
i	0	88.89	5.56	0	0	0	2.78	0	2.78
e	0	0	63.89	27.78	0	0	0	0	8.33
æ	0	0	19.44	75	5.56	0	0	0	0
a	0	0	0	3.33	56.67	40	0	0	0
o	0	0	0	0	40.63	46.88	9.38	0	3.13
u	0	0	4.55	0	4.55	0	50	0	40.91
uw	0	0	0	0	0	0	26.67	73.33	0
X	0	0	14.71	5.88	14.71	26.47	0	0	38.24

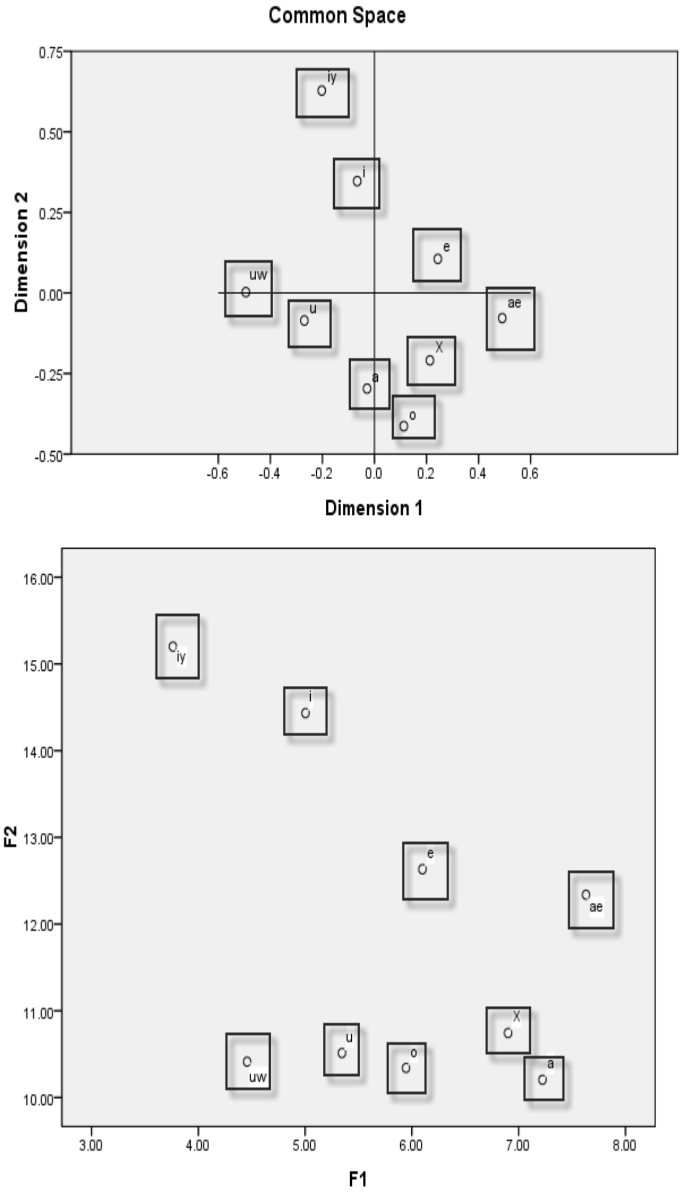
The first column shows stimuli while the first row shows responses. And each of the cells has an identification rate. For example, when /iy/ stimuli were presented, listener1 correctly identified them 80.56%. However, s/he incorrectly identified them as /i/ 13.89%, as /u/ 2.78%, and finally as /uw/

2.78%. This is an asymmetric matrix and will be converted into a symmetric matrix with Pearson correlation across responses, to be fed onto an MDS analysis.

The identification accuracy rate for each vowel in the diagonal tells us how well a listener can identify it. A high misidentification rate in each of the other cells, means serious confusion between a stimulus and a response vowel. A confusion matrix shows relative perceptual similarity between a target vowel and the other vowels through correct or wrong responses. Misidentification rates in cells in the same row are also important, as they offer vowel confusability information. An MDS analysis converts vowel misidentification rates into the perceptual similarity relations between two AE vowels on a space. As a result, vowel points are visually represented on a perceptual space.

An MDS analysis (PROXSCAL, weighted Euclidean, multiple matrix sources, weighted, similarity proximities, linear combination of independent variables) conducted. Two-dimensional coordinates could be fit to the 9 target AE vowels to match these inter-response rate distances with relatively little stress (normalized stress = 0.059), which means that the distances between the spatial coordinates given to each response variables satisfiably corresponded inversely to the identification similarities generated from the confusion matrices.

The derived perceptual map (top) below shows that the diagonal may be interpreted as correlating with height and the anti-diagonal with roundness and/or frontness. More importantly, all AE vowels are spaced evenly on the map; they are relatively well identified from each other by AE listeners. All the AE vowels are almost equally spaced in distance. On the other hand, an acoustically derived AE vowel space with scatterplots of F2 vs. F1 in bark of the stimuli spoken by AE talkers is shown in the bottom figure. If the perceptual map (top) was tilted 45 degrees clockwise, the two maps with vowel points look very similar.



Figures 6. Perceptual vowel space for two AE listeners (top) and acoustic vowel space for the stimuli (bottom)

They show almost the same vowel chart shape: a trapezoid shape. The perceptual vowel space (top) shows almost the same spacing between neighboring vowels as the one based on the acoustic data. The only difference between the two lies in the plots of /o/ and /a/. /a/ and /o/ are switched over from each other. However, when only perceptual confusion between perceptually similar vowels is focused in this study, this error is tolerable. Other than that, the two maps are almost equivalent.

For a further reference, the acoustic map in Hillenbrand et al. (1995) is also considered. They collected the data from 139 Southern Michigan talkers (45 men, 48 women, and 46 children) and data from P&B (1952) and converted the formant values to a Bark scale. The F1/F2 frequencies were plotted.

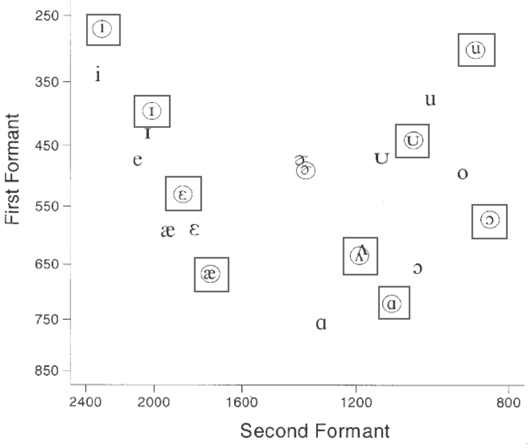


Figure 7. A mirror image of f1/f2 frequency plots in bark for adult male talkers from Hillenbrand et al. (1995; southern Michigan) and from P&B (1952; mid-Atlantic). The symbols in rectangles are from P&B (1995).

The acoustic data in P&B (1952) are largely from mid-Atlantic talkers. Perceived mean vowel points are represented as symbols in rectangles. Note that the un-circled vowels are from southern Michigan talkers. For our purpose, however, only the circled vowels are considered, since southern Michigan English is a seriously accented version of AE. The perceptual vowel space for AE listeners in our work patterns together with F1/F2 frequency-based vowel plots from the data in P&B. The formant values were measured at a time when the formant pattern was maximally steady, based on visual inspection of a spectrogram (N = 33 for the P&B data). The vowel points in rectangles in figure 9 patterns exactly together with the two sets of the vowel points in figure 8.

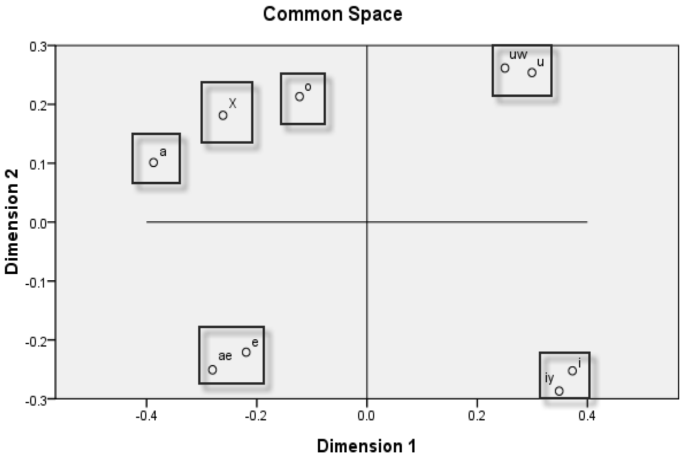
We will show that the perceptual map can also simulate the performance of Korean listeners as to how well or badly they perceive perceptually similar AE vowels.

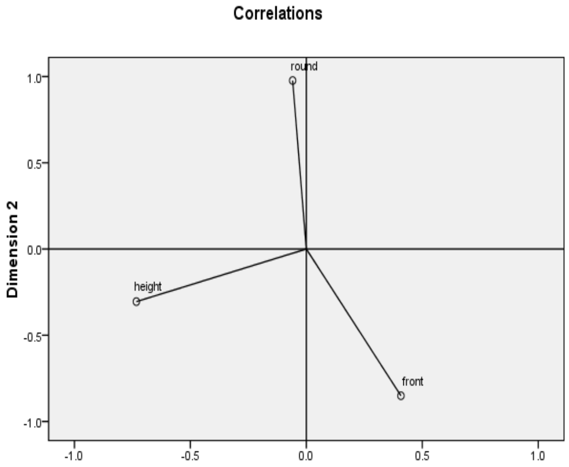
3.3 Comparison between perceptual vowel space for Korean listeners and acoustic vowel space

Thirty-four vowel confusion matrices for nine AE vowels from all listeners were obtained from the identification test. They were converted to Pearson correlation matrix across responses and then fed onto an MDS analysis.

To scale the model, Euclidean was weighted and ratio proximity transformations were applied within each source separately. Common space was further restricted by linearly regressing independent phonological vowel feature variables: height, roundness, and backness. These independent phonological feature variables were transformed individually across the vowel responses (Optimal scaling level: ordinal with ties allowed to be untied). Then the transformed independent variables were regressed on the dimensions of the common space and correlations between the transformed independent variables and two optimal dimensions were taken, based on the stress (Kruskal et al. 1978).

Two-dimensional coordinates could be fit to the 9 target AE vowels to match these inter-response rate distances with a very little stress (normalized stress = 0.013).

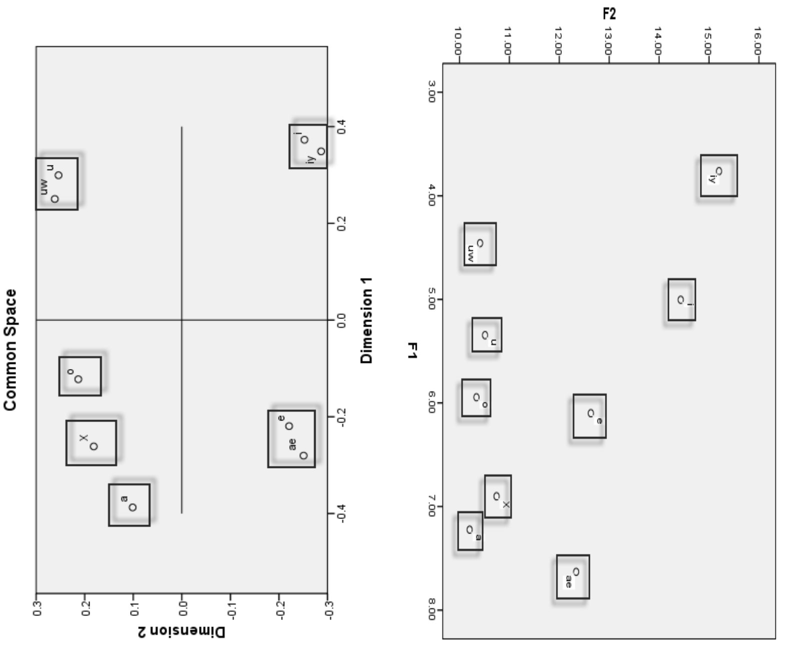




Figures 8. Perceptual vowel space for Korean listeners (top) and feature correlation map (bottom)

The resulting transformed feature variable correlation on the same dimension coordinates (bottom) suggests that dimension 1 may be interpreted more or less as height in the perceptual vowel space (top) while dimension 2 as roundness.

In the perceptual vowel space, three remotely separate clusters are formed: /iy, i/, /u, uw/, and /e, æ/. The two members of each of the three vowel pairs were represented as two vowel points next to each other, marked with a rectangle. Clearly, Korean listeners' perception patterns for AE vowels are seriously distorted except for AE /o/, /a/, and /X/, which are represented as separate points. This means that Korean listeners had identification problems between /iy/ and /i/, between /u/ and /uw/, and between /e/ and /æ/. However, /o/, /a/, and /X/ were relatively better identified, though still bad.



Figures 9. A perceptual AE vowel space for Korean listeners (left) and an acoustic vowel space for the stimuli (right)

When the perceptual AE vowel space for Korean listeners (left) is compared with the acoustic AE vowel space for the stimuli spoken by AE talkers (right), it is found that the three vowel clusters with two members in on the perceptual space (left) are the three perceptually difficult vowel pairs for Korean listeners. Namely, when AE /uw/ and /u/ were presented to Korean listeners, they seemed to be mostly identified as Korean /u/. When AE /e/ and /æ/ were heard, they seemed to be mostly identified as Korean /e/. Lastly, when AE /iy/ and /i/ were presented, they were mostly identified as Korean /i/. And these three pairs form three separate clusters on the perceptual space.

We waged another experiment in which Korean listeners were evaluated on perception of AE vowels in the same phonetic environment: hVd. This time, however, the vowels were not manipulated. This is because the result will represent the listener's perceptual ability better than that of the previous experiment in which the end of the vowel portion of the stimulus was trimmed off to avoid potential lexical effects. The result will be fed into an MDS analysis on an individual basis.

4. Experiment II: Identification test without manipulation and with hVd stimuli

4.1 Subjects

The subjects were 10 university-level Korean students (5 males and 5 females) ranging in age from 20 to 25 years old. These listeners were the ones who did not take part in experiment I. All of them had at least 6 years of prior English instruction at the middle and high school levels. None of the subjects had any reported history of speech or hearing problems.

4.2 Stimuli

The objective of this hVd test was to force the Korean listeners to identify the correct vowels in response to the presentations of the English hVd speech stimulus. The nine target English vowels were primary-stressed in hVd: /iy, i, e, æ, a, o, u, uw, X(ɐ)/. 17 different males, females, boys and girls produced 9 sample words, totaling 153 tokens, as shown below (17 speakers * 9 vowels = 153 tokens) (data from Hillenbrand et al. 1995).

Table 5. hVd matrix of stimuli

	iy	I	e	æ	a	o	u	uw	X	total
p	heed	hid	head	had	hod	hawed	hood	who'd	hud	9
total	17	17	17	17	17	17	17	17	17	153

4.3 Procedure

This hVd perception test was done through the presentation of the randomized 153 hVd tokens. A brief introduction to 9 AE vowels and their respective transcriptions was given to the listeners before the test. Then a software module based on Alvin (Hillenbrand et al. 2005), provided a computer screen showing the 9 vowel icons with the sample words. Each of the vowel icons was filled with an IPA symbol and the stimulus word: *heed*, *hid*, *head*, *had*, *hod*, *hawed*, *hood*, *who'd*, and *hud*. No problem was reported during the experiment since subjects were already familiar with the example words and IPA symbols.

Each listener heard the randomized stimuli presented via a PC over a headphone, and was forced to click on one of the 9 vowel icons for each presentation s/he heard. In order to make the loudness consistent, the sound volume was fixed to a comfortable listening level by the author, and listeners could not adjust the volume without the author's permission. After the click for vowel identification, there was a pause of 400ms before the next stimulus was presented. Without clicking, the next stimulus was not presented. When a listener made an error click, s/he could go back and make a readjustment click after listening to the previous presentation again. S/he could listen to each presentation repeated up to 3 times. The identification experiment for each listener took about less than 10 minutes on average. This test setup allowed participants to proceed at their own comfortable pace.

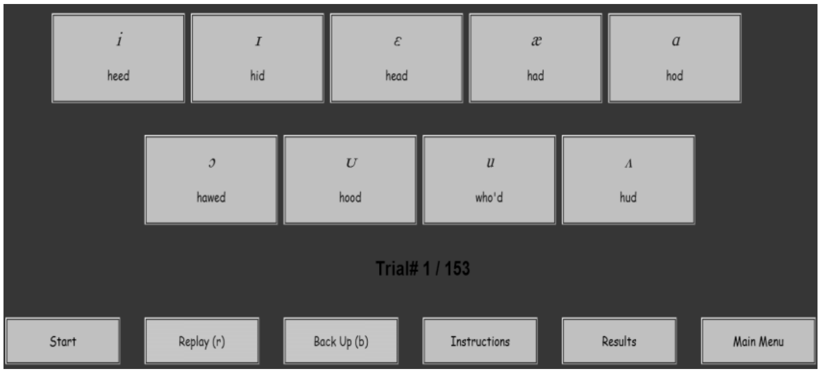


Figure 10. Computer screen for hVd identification test

4.4 Models

The resulting confusion matrix for listener1 was converted into a Pearson correlation matrix with variable distances computed for proximities and then was fed into a PROXSCAL MDS analysis in SPSS:

Table 6. Pearson correlation matrix for listener1 with variable distances computed for similarities

	iy	I	e	æ	a	o	u	uw	X
iy	1.000	.686	-.249	-.272	-.275	-.335	-.205	-.126	-.314
I	.686	1.000	-.182	-.254	-.257	-.313	-.193	-.196	-.293
e	-.249	-.182	1.000	.619	-.199	-.304	-.292	-.222	-.310
æ	-.272	-.254	.619	1.000	-.261	-.311	-.279	-.208	-.218
a	-.275	-.257	-.199	-.261	1.000	.685	-.283	-.211	.243
o	-.335	-.313	-.304	-.311	.685	1.000	-.233	-.241	.417
u	-.205	-.193	-.292	-.279	-.283	-.233	1.000	.517	-.242
uw	-.126	-.196	-.222	-.208	-.211	-.241	.517	1.000	-.229
X	-.314	-.293	-.310	-.218	.243	.417	-.242	-.229	1.000

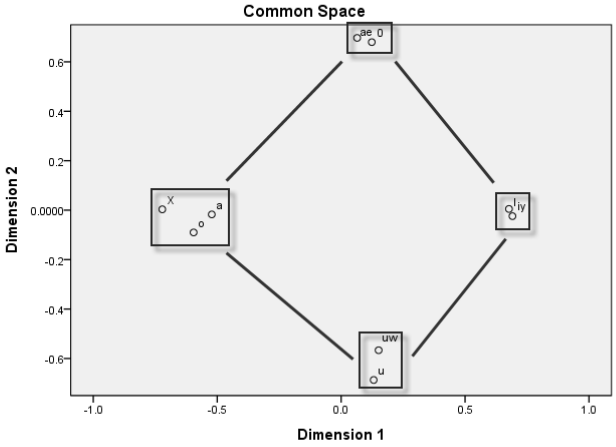


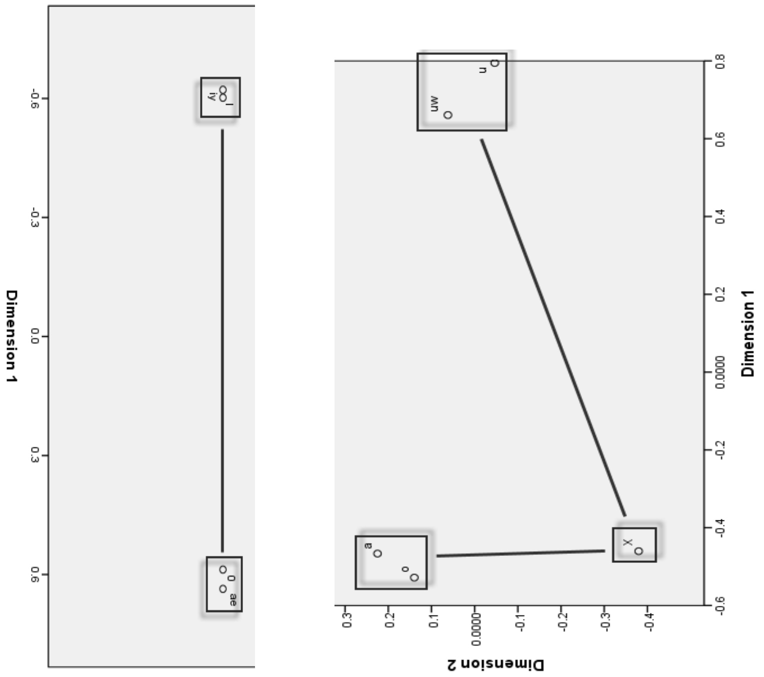
Figure 11. The perceptual vowel space for listener1

The 2-dimensional perceptual vowel map is for Korean listener1 (normalized stress: 0.0267). She had a more serious problem with the perceptual distinction between [iy] and [i] (upper left-hand) and between [uw] and [u] (upper right-hand) than between [e][5] and [æ] (lower right-hand). However, the perceptual confusability between [a], [o], and [X] turned out to be unclear. Despite the low stress, the resulting vowel space for listener1 is not satisfactory. I will call this analysis “single matrix model”, since only one matrix is fed into an MDS analysis

We will propose a more optimal model (“the split matrix model”) to reduce stress by manipulating the confusion matrix before being fed into an MDS analysis. We split the confusion matrix with all nine vowels into two separate confusion matrices of front vowels and back vowels and then fed them into an MDS analysis separately. As a result, we will get two perceptual vowel maps for the two vowel groups. This is possible because we observed that most Korean listeners were not confused between front and back vowels. This strategy can drastically reduce the overall stress level.

The resulting figures below after MDS analyses, show that front vowels can be optimally represented on a 1-dimensional map (normalized stress: 0.0007). This is interpreted as only one dimension (vowel height) is enough to represent all the front vowels. On the other hand, back vowels are best represented on a 2-dimensional map (normalized stress: 0.0012). Listener1’s perception is represented below by the split matrix model:

[5] Due to an unknown bug with SPSS 18, [e] next to [æ] is represented as [0]. This bug also occurs in figures 14.



Figures 12. Perceptual vowel space for listener1 by the split matrix model

When the two maps are positioned side by side, a complete perceptual vowel space for the all nine vowels can be derived. The resulting vowel space better represents listener1's perception. Listener1 has a relatively less serious problem in differentiating between [uw] and [u] compared to between the [iy] and [i] and between [e] and [æ]. Furthermore, she experiences severe difficulty differentiating between [a] and [o], which was not fully represented in the previous model. Since the less is stress, the better the model turns out, this split matrix model better explains the perceptual vowel space for listener1.

5. Conclusion

Korean speakers' AE vowel production can be easily visualized on a two-dimensional acoustic vowel map through measuring F1 and F2 values. As a result, it is easy to interpret which vowels a given speaker has problems with, since the problematic vowels will be represented close to each other on a two-dimensional acoustic map. However, Korean speakers' perceptual abilities of AE vowels have not been easy to evaluate. The absolute accuracy rate in a listener's perception of a given vowel alone is not a good measure for evaluation. A listener's relative perceptual confusability between vowels also has to be considered for a better and more precise evaluation. Confusion matrices from identification tests are a major source for further evaluation. However, they are extremely difficult to interpret. As demonstrated in this paper, an MDS analysis of the resulting matrices from identification tests (or discrimination tests) is a good tool to visualize the perceptual proximities between AE vowels felt by Korean listeners. The MDS approach in this paper crucially utilizes listeners' perception errors between perceptually similar vowels, which are visually realized as proximities between vowel points on the perceptual vowel space.

This paper proposed an optimized MDS model, the split matrix model, which uses split front and back vowel matrices. This model may drastically reduce stress and hence represent listeners' perceptual space far more accurately. However, further research is necessary to refine the analysis since the model collapses when Korean listeners make identification errors between front [æ] and back [a], though such errors were rarely observed. More researches on optimization processes are necessary for more refined models.

REFERENCES

BEST, CATHERINE T. 1995. A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, 171-204. Baltimore, MD: York Press.

BEST, CATHERINE T., GERALD W. MCROBERTS, and NOMATHEMBA M. SITHOLE. 1988. Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English speaking adults and infants. *Journal of Experimental Psychology* 14, 345-360.

CARROLL, DOUGLAS J. and JIH-JIE JANG. 1970. Analysis of individual differences in multidimensional scaling via n-way generalization of "Eckart-Young" decomposition. *Psycometrika* 35, 283-319.

Dong-A's Prime Dictionary. 1997. Seoul: Doosan Dong-A.

E4u Dictionary. 2001. Seoul: YBMSisa.

FLEGE, JAMES E. 1988. The production and perception of speech sounds in a foreign language. In H. Winitz (ed.). *Human Communication and Its Disorders, A Review*, 224-401. Norwood, NJ: Ablex.

FLEGE, JAMES E. 1995. Second language speech learning: Theory, findings and problems. In W. Strange (ed.). *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues*, 233-277. Timonium, MD: York Press.

FLEGE, JAMES E., MURRAY J. MUNRO, and ROBERT A. FOX. 1994. *JASA* 95.6, 3623-3641.

FOX, ROBERT A. 1983. Perceptual structure of monophthongs and diphthongs in English. *Language and Speech* 26, 21-60.

FOX, ROBERT A., JAMES E. FLEGE, and MURRAY J. MUNRO. 1995. The perception of English and Spanish vowels by native English and Spanish listeners: A multidimensional scaling analysis. *JASA* 97.4, 2540-2551.

HARNSBERGER, JAMES D., MARIO A. SVIRSKY, ADAM R. KAISER, DAVID B. PISONI, RICHARD WRIGHT, and TED A. MEYER. 2001. Perceptual "vowel spaces" of cochlear implant users: Implications for the study of auditory adaptation to spectral shift. *JASA* 109.5, 213555-2145.

HILLENBRAND, JAMES M. and ROBERT T. GAYVERT. 2005. Open source software for experiment design and control. *Journal of Speech, Language, and Hearing Research* 48, 45-60.

HILLENBRAND, JAMES D., LAURA A. GETTY, MICHAEL J. CLARK, and KIMBERLEE WHEELER. 1995. Acoustic characteristics of American English vowels. *JASA* 97.5, 3099-3111.

HONG, SOONHYUN. 2007a. The characteristics of vowel identification errors of university-level Korean students of American English: HCA. *Language and Linguistics* 39, 257-277. Language Research Institute, Hankuk University of Foreign Studies.

HONG, SOONHYUN. 2007b. The effects of V-features on American English vowel perception by university-level Korean talkers. *Studies in Modern Grammar* 48, 145-170.

HONG, SOONHYUN. 2009. Training Korean listeners to perceive American English fricatives and affricates: A listener-customized adaptive approach. *Studies in Phonetics, Phonology and Morphology* 15.1, 147-170.

JOHNSON, KEITH and JOAN MULLENNIX. 1977. Complex representations used in speech processing: overview of the book. In J., Keith. and J. Mullennix (eds.), *Talker Variability in Speech Processing*, 1-8. San Diego, CA: Academic Press.

KLIEN, WILLIAM, REINIER PLOMP, and LOUIS C. POLS. 1970. Vowel spectra, vowel spaces, and vowel identification. *JASA* 48.4, 999-1009.

KRUSKAL, JOSEPH B. and MYRON WISH. 1978. *Multidimensional Scaling*. Beverly Hills and London: Sage Publications.

LADEFOGED, PETER. 2006. *A Course in Phonetics*, 5th edition. Boston, MA: Thomson.

LAMBACHER, STEPHEN G., WILLIAM L. MARTENS, and GARRY MOLHOLT. 2000. A comparison of identification of American English vowels by native speakers of Japanese and English. In H. Umeda (ed.), *Proceedings of the Phonetic Society of Japan Meeting*, 213-218. Tokyo, Japan: Phonetic Society of Japan.

LIBERMAN, ALVIN 1957. Some results of research on speech perception. *JASA* 29, 117-123.

NISHI, KANAE and DIANE KEWLEY-PORT. 2007. Training Japanese listeners to perceive American English vowels: Influence of training sets.

Journal of Speech, Language, and Hearing Research 50, 1496-1509.

PETERSON, GORDON E. and HAROLD L. BARNEY. 1952. Control methods used in a study of the vowels. *JASA* 24.2, 175-184.

POLS, LOUIS C. W., LEO J. TH. VAN DER KAMP, and REINIER PLOMP. 1969. Perceptual and physical space of vowel sounds. *JASA* 46.2, 458-467.

RAKERD, BRAD and ROBERT R. VERBRUGGE. 1985. Linguistic and acoustic correlates of the perceptual structure found in an individual differences scaling study of vowels. *JASA* 77.1, 296-301.

SHEPARD, ROGER N. 1972. Psychological representation of speech sounds. In E. David and P. Denes (eds.), *Human Communication: A Unified View*. New York: McGraw Hill.

SINGH, SADANAND and DAVID R. WOODS. 1970. Perceptual structure of 12 American English Vowels. *JASA* 49.6, 1861-1866.

STEVENS, KENNETH, ALVIN LIBERMAN, MICHAEL STUDDERT-KENNEDY, and SVEN ÖHMAN. 1969. Cross language study of vowel perception. *Language and Speech* 12, 1-23.

STRANGE, WINIFRED, JAMES J. JENKINS, and THOMAS L. JOHNSON. 1983. Dynamic specification of coarticulated vowels. *JASA* 74.3, 695-705.

STRANGE, WINIFRED, REIKO AKANE-YAMADA, RIEKO KUBO, SONJA A. TRENT, KANAE NISHI, and JAMES J. JENKINS. 1998. Perceptual assimilations of American English vowels by Japanese listeners. *Journal of Phonetics* 26, 311-344.

TERBEEK, DALE 1977. A cross-language multi-dimensional scaling study of vowel perception. *Working Papers in Phonetics* 37. UCLA.

VERBRUGGE, ROBERT and WINIFRED STRANGE. 1976. What information enables a listener to map a talker's vowel space? *JASA* 60.1, 198-212.

YAMADA TSUNEO, REIKO YAMADA, and WINIFRED STRANGE. 1995. Perception of English vowels and consonants by Japanese learners of English. In *Proceedings of the Acoustical Society of Japan*, 379-380. Utsunomiya, Japan: Acoustical Society of Japan.

Soonhyun Hong
Department of English Language and Literature
Inha University
253 Yonghyun-dong, Nam-gu, Incheon, Korea
e-mail: shong@inha.ac.kr

received: March 4, 2011
accepted: April 15, 2011